Inequality, Probability, and Joviality

- In many cases, we don't know the true form of a probability distribution
  - E.g., Midterm scores
  - But, we know the mean
  - May also have other measures/properties
    - Variance
    - Non-negativity
    - Etc.
  - Inequalities and bounds still allow us to say something about the probability distribution in such cases
    - May be imprecise compared to knowing true distribution!

Markov's Inequality

- Say X is a non-negative random variable

$$P(X \geq a) \leq \frac{E[X]}{a}, \quad \text{for all } a > 0$$

- Proof:
  - I = 1 if X ≥ a, 0 otherwise
  - Since $X \geq 0, \quad I \leq \frac{X}{a}$
  - Taking expectations:

$$E[I] = P(X \geq a) \leq E\left[\frac{X}{a}\right] = \frac{E[X]}{a}$$

Andrey Andreyevich Markov

- Andrey Andreyevich Markov (1856-1922) was a Russian mathematician

  - Markov's Inequality is named after him
  - He also invented Markov Chains…
    - …which are the basis for Google's PageRank algorithm
  - His facial hair inspires fear in Charlie Sheen

Markov and the Midterm

- Statistics from last quarter's CS109 midterm
  - X = midterm score
  - Using sample mean $\overline{X}$ = 78.1 ≈ E[X]
  - What is P(X ≥ 91)?

$$P(X \geq 91) \leq \frac{E[X]}{91} = \frac{78.1}{91} \approx 0.8582$$

  - Markov bound: ≤ 85.82% of class scored 91 or greater
  - In fact, 34.44% of class scored 91 or greater
    - Markov inequality can be a very loose bound
    - But, it made <u>no</u> assumption at all about form of distribution!

Chebyshev's Inequality

- X is a random variable with E[X] = μ, Var(X) = σ^2

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}, \quad \text{for all } k > 0$$

- Proof:
  - Since $(X - \mu)^2$ is non-negative random variable, apply Markov's Inequality with $a = k^2$

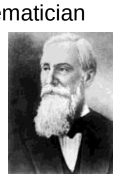
$$P((X - \mu)^2 \geq k^2) \leq \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

  - Note that: $(X - \mu)^2 \geq k^2 \Leftrightarrow |X - \mu| \geq k$, yielding:

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Pafnuty Chebyshev

- Pafnuty Lvovich Chebyshev (1821-1894) was also a Russian mathematician



  - Chebyshev's Inequality is named after him
    - But actually formulated by his colleague Irénée-Jules Bienaymé
  - He was Markov's doctoral advisor
    - And sometimes credited with first deriving Markov's Inequality
  - There is a crater on the moon named in his honor

Of the Midterm What Say You Chebyshev?

- Statistics from last quarter's CS109 midterm
  - X = midterm score
  - Using sample mean $\overline{X} = 78.1 \approx E[X]$
  - Using sample variance $S^2 = (24.5)^2 = 600.25 \approx \sigma^2$
  - What is P(| X – 78.1 | ≥ 30)?

$$P(|X - E[X]| \geq 30) \leq \frac{\sigma^2}{(30)^2} = \frac{600.25}{900} \approx 0.6669$$

$$P(|X - E[X]| < 30) = 1 - P(|X - E[X]| \geq 30) \geq 1 - 0.6669 = 0.3331$$

  - Chebyshev bound: ≤ 66.69% scored ≥ 108.1 or ≤ 48.1
  - In fact, 21.85% of class scored ≥ 108.1 or ≤ 48.1
    - Chebyshev's inequality is really a theoretical tool

One-Sided Chebyshev's Inequality

- X is a random variable with E[X] = 0, Var(X) = σ^2

$$P(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}, \quad \text{for any } a > 0$$

  - Equivalently, when E[Y] = μ and Var(Y) = σ^2:

$$P(Y \geq E[Y] + a) \leq \frac{\sigma^2}{\sigma^2 + a^2}, \quad \text{for any } a > 0$$

$$P(Y \leq E[Y] - a) \leq \frac{\sigma^2}{\sigma^2 + a^2}, \quad \text{for any } a > 0$$

  - Follows directly by setting X = Y – E[Y], noting E[X] = 0

Comments on Midterm, One-Sided One?

- Statistics from last quarter's CS109 midterm
  - X = midterm score
  - Using sample mean $\overline{X} = 78.1 \approx E[X]$
  - Using sample variance $S^2 = (24.5)^2 = 600.25 \approx \sigma^2$
  - What is P(X ≥ 103.1)?

$$P(X \geq 78.1 + 25) \leq \frac{600.25}{600.25 + (25)^2} \approx 0.4899$$

  - One-sided Chebyshev bound: ≤ 48.99% scored ≥ 103.1
  - In fact, 13.26% of class scored ≥ 103.1
  - Using Markov's inequality: $P(X \geq 103.1) \leq \dfrac{78.1}{103.1} \approx 0.7575$

Chernoff Bound

- Say we have MGF, M(*t*), for a random variable X
  - Chernoff bounds:

$$P(X \geq a) \leq e^{-ta} M(t), \quad \text{for all } t > 0$$
$$P(X \leq a) \leq e^{-ta} M(t), \quad \text{for all } t < 0$$

  - Bounds hold for $t \neq 0$, so use *t* that minimizes e^{-ta}M(*t*)
- Proof:
  - X has MGF: M(*t*) = E[e^{tX}]
  - Note P(X ≥ *a*) = P($e^{tX} \geq e^{ta}$), use Markov's inequality:

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E[e^{tX}]}{e^{ta}} = e^{-ta} E[e^{tX}] = e^{-ta} M(t), \quad \text{for all } t > 0$$

  - Similarity for P(X ≤ a) when *t* < 0

Herman Chernoff

- Herman Chernoff (1923-) is an American mathematician and statistician

  - Chernoff Bound is named after him
    - And it actually was derived by him!
  - He is Professor Emeritus of Applied Mathematics at MIT and of Statistics at Harvard University
    - I do not know if he is a fan of Charlie Sheen

Chernoff's Feeling (Unit) Normal

- Z is standard normal random variable: Z ~ N(0, 1)
  - Moment generating function: $M_Z(t) = e^{t^2/2}$
  - Chernoff bounds for P(Z ≥ *a*)

$$P(Z \geq a) \leq e^{-ta} e^{t^2/2} = e^{t^2/2 - ta}, \quad \text{for all } t > 0$$

  - To minimize bound, minimize: $t^2/2 - ta$
    - Differentiate w.r.t. *t*, and set to 0: $t - a = 0 \Rightarrow t = a$

$$P(Z \geq a) \leq e^{-a^2/2}, \quad \text{for all } t = a > 0$$

  - Can proceed similarly for *t* = *a* < 0 to obtain:

$$P(Z \leq a) \leq e^{-a^2/2}, \quad \text{for all } t = a < 0$$

  - Compare to: $P(Z > z) = 1 - P(Z \leq z) = 1 - \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$

Chernoff's Poisson Pill

- X is Poisson random variable: X ~ Poi(λ)
  - Moment generating function: $M_X(t) = e^{\lambda(e^t - 1)}$
  - Chernoff bounds for P(X ≥ *i*)

$$P(X \ge i) \le e^{\lambda(e^t-1)} e^{-it} = e^{\lambda(e^t-1)-it}, \quad \text{for all } t > 0$$

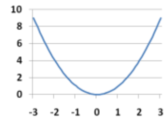
  - To minimize bound, minimize: $\lambda(e^t - 1) - it$
    - Differentiate w.r.t. *t*, and set to 0: $\lambda e^t - i = 0 \Rightarrow e^t = i/\lambda$

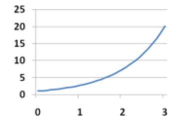
$$P(X \ge i) \le e^{\lambda(i/\lambda - 1)} \left(\frac{i}{\lambda}\right)^{-i} = e^i e^{-\lambda} \left(\frac{\lambda}{i}\right)^i = \left(\frac{e\lambda}{i}\right)^i e^{-\lambda}, \quad \text{for all } i/\lambda > 1$$

  - Compare to: $P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}$

Jensen's Inequality

- If *f*(*x*) is a *convex* function then E[*f*(*x*)] ≥ *f*(E[X])
  - *f*(*x*) is **convex** if $f''(x) \ge 0$ for all *x*
  - Intuition: Convex = "bowl". E.g.: $f(x) = x^2$, $f(x) = e^x$





  - if *g*(*x*) = -*f*(*x*) is convex, then *f*(*x*) is **concave**
  - Proof outline: Taylor series of *f*(*x*) about μ. Be happy.
  - Note: E[*f*(*x*)] = *f*(E[X]) only holds when *f*(*x*) is a line
    - That is when: $f''(x) = 0$ for all *x*

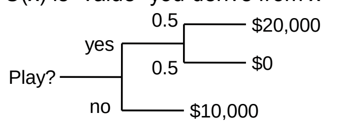
Johan Jensen

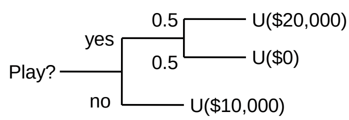
- Johan Ludwig William Valdemar Jensen (1859-1925) was a Danish mathematician

  - He derived Jensen's inequality
  - He was president of the Danish Mathematical Society from 1892 to 1903
  - He has more names than Charlie Sheen

A Brief Digression on Utility Theory

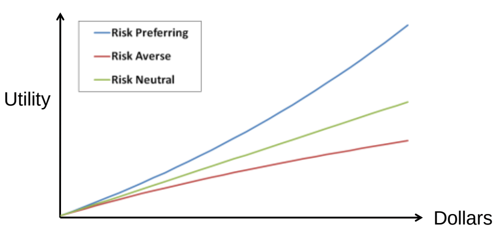
- Utility U(x) is "value" you derive from x





  - Can be monetary, but often includes intangibles
    - E.g., quality of life, life expectancy, personal beliefs, etc.

Utility Curves



- Utility curve determines your "risk preference"
  - Can be different in different parts of the curve
  - We'll talk more about this near the end of the quarter

Jensen's Investment Advice

- Example: risk-taking investor, with two choices:
  - Choice 1: Invest money to get return X where E[X] = μ
  - Choice 2: Invest money to get return μ (probability 1)
- Want to maximize utility: *u*(R), where R is return
  - if *u*(X) convex then E[*u*(X)] ≥ *u*(μ), so choice 1 better
  - If *u*(X) concave then E[*u*(X)] ≤ *u*(μ) so choice 2 better
  - Convex *u* ⇒ "risk preferring", concave *u* ⇒ "risk averse"